What is claimed is:

1. A method to predict the topology of the spatial arrangement of an amino acid sequence using an entropy evaluation model that takes into account the global contributions of entropy to the folding of a biopolymer (herein referred to by the name cross linking entropy (CLE) and described in the literature) combined with other thermodynamic potentials as a protein-folding model.

2. A method according to claim 1, comprising the following steps:

A. inputting an amino acid sequence of a protein,

B. preparing information on the secondary structure of the said amino acid sequence by way of at least one theoretical or experimental estimate,

C. applying the CLE method to the said amino acid sequence and secondary structure information to evaluate the free energy of a combinatorial number of β-strand and α-helix arrangements as rapidly as polynomial time: $c(n-1)(n+1)$ wherein c is a constant and n is the number of secondary structure elements found in the said amino acid in 2A and prepared in 2B,

D. applying the CLE method in conjunction with other thermodynamic potentials that approximate hydrophobic, electrostatic and polar interactions, but not limited to these aforementioned thermodynamic potentials stated

herein, in a thermodynamic calculation to account for both short and long range folding interactions and predict a minimum free energy and corresponding topology of the said amino acid sequence,

E. applying the CLE method to predict the global folding kinetics of the said amino acid sequence, and

F. storing the information in a data file or in other form of digital memory.

3. A method according to claim 1 and 2, in which the cross linking entropy (CLE), which is an entropy evaluation model that takes into account the global effects of entropy in the folding of a biopolymer, is used to evaluate the entropy loss of a protein due to folding into a particular topology given a known secondary or estimated secondary structure.

4. A method according to claim 3, in which loss of biological activity of the protein can be further predicted.

5. A method according to claim 4, in which a initial theoretical estimate of the secondary structure is obtained from either a theoretical source, an experimental source such as an NMR experiment or x-ray crystallography, or both.

6. A method according to claim 5, in which the theoretical estimate is further supplemented with sequence alignment to find regions in which conserved segments remains essentially unchanged by differences in the aligned sequences.

7. A method according to claims 5 and 6 in which the said amino acid sequence and secondary structure information is used to evaluate the free energy of a combinatorial number of β-strand and α-helix arrangements as rapidly as polynomial time: $c(n-1)(n+1)$ wherein c is a constant and n is the number of secondary structure elements found in the said amino acid and obtained.

8. A method to predict the topology of the spatial arrangement of an amino acid sequence comprising following steps:

A. inputting an amino acid sequence of a protein,

B. preparing information on the secondary structure of the said amino acid sequence by way of at least one theoretical or experimental estimate,

E. applying the CLE method to approximate the global folding kinetics of the said amino acid sequence,

G. applying the CLE method to the said amino acid sequence and secondary structure information to reduce the combinatorial number of β-strand and α-helix arrangements to a computationally manageable number, and

H. applying the CLE method to optimize the free energy to find the most thermodynamically favorable topology for the said amino acid sequence,-

wherein the global free energy (FE) contribution from the CLE between two distinct amino acid residues, herein labeled *i*

and j, is calculated by equation (1):

$$\Delta G_{ij} = -T\Delta S_{ij} = \frac{\gamma k_B T}{\xi}\left\{\ln\left(\frac{2\gamma\xi\Delta N_{ij}}{3\lambda_{ij}^2}\right) - 1 + \frac{3\lambda_{ij}^2}{2\gamma\xi\Delta N_{ij}}\right\} \quad (1)$$

wherein, i and j represent the indices of two distinct residues in the said amino acid sequence, and $j > i$, $\Delta N_{ij} = j - i + 1$ expresses the number of residues separating i and j, ΔG_{ij} is the difference in the free energy contribution to the CLE from residues i and j transitioning from the denatured (random flight) state to the native state, ΔS_{ij} is the corresponding entropy loss, ξ is the persistence length, γ is a dimensionless weight parameter describing the self-avoiding properties of a polymer chain, k_B is the Boltzmann constant, T is the temperature, and λ_{ij} (the bond gap) expresses the amino acid separation distance between the center of mass of residue i and the center of mass of residue j when both are treated as isolated molecules.

9. A method according to claim 8, in which the total CLE contribution to the free energy (ΔG_{cle}) is calculated by equation (2):

$$\Delta G_{cle} = \Delta G_{\xi}^{o} + \sum_{all\_bonds(i,j)} \Delta G_{ij} + \sum_{i',j'} f_{i'j'}(\xi) \quad (2)$$

wherein, ΔG_{ij} is defined in equation (1), i' and j' are indices specifying two secondary structure elements (α-helices or β-strands) that are joined together by the corresponding set of bonds i and j, $f_{i'j'}(\xi)$ is a positive definite penalty function used to enforce topology constraints on the minimum allowed sequence length of a loop connecting two elements of secondary structure i'-j' and is a function of the persistence length ξ, and ΔG^{o}_{ξ} is a renormalization correction and is an integral function of ξ as shown by equation (3):

$$\Delta G^{o}_{\xi} = \frac{(\gamma + 1/2)Nk_BT}{D\xi}\int_{1}^{\xi}\left(\frac{\ln(x)}{(1-x)}+1\right)dx \qquad (3)$$

wherein, ξ, γ, k_B, and T mean the same as defined in claim 7, N indicates the number of amino acids in the said sequence, D is the dimensionality of the system, the limits in the integral $(1 \rightarrow \xi)$ indicate the change in the number of degrees of freedom from an individual amino acid reside to a cluster of ξ amino acids treated as a group (where ξ>1 amino acid and ξ need not be an integer) and x is dummy variable in the integral substituting for ξ.

10. A method according to claim 9, in which the optimal β-sheet alignments are obtained by using thermodynamics.

11. A method according to claim 8 to 10, in which the CLE method is applied in conjunction with other derived or constructed

thermodynamic potentials that approximate hydrophobic, electrostatic and polar interactions, in a thermodynamic calculation to account for both short and long range folding interactions and predict a minimum free energy and corresponding topology of the said amino acid sequence.

12. A method for building a 3D structure of a protein for MD simulation from the topology obtained by the method according to any one of claims 1,2 and 4-10.

13. A method according to claim 1, comprising the following steps:

A. obtaining an amino acid sequence of a protein,

B. preparing information on the secondary structure of the said amino acid sequence by way of at least one theoretical or experimental estimate,

E. applying the CLE method to approximate the global folding kinetics of the said amino acid sequence,

I. using the global folding kinetics to predict the optimal topology of the said amino acid sequence, and

F. storing the information in a data file or in other form of digital memory.